



# **STIC Search Report**

## **EIC 2100**

**STIC Database Tracking Number: 1944**

**TO: Thanh-ha Dang**  
**Location: RND 3B15**  
**Art Unit: 2163**  
**Friday, June 30, 2006**

**Case Serial Number: 09/780123**

**From: Lucy Park**  
**Location: EIC 2100**  
**RND-4B11**  
**Phone: 571-272-8667**

**lucy.park@uspto.gov**

### **Search Notes**

Dear Examiner Dang,

Here are the search results for your Fast & Focused search request on case number 09/780123. I flagged the results that looked most relevant, but please review all of the results. Please let me know if you have any questions about these or if you need any further information.

Lucy

File 347:JAPIO Dec 1976-2005/Dec(Updated 060404)  
(c) 2006 JPO & JAPIO  
File 350:Derwent WPIX 1963-2006/UD,UM &UP=200641  
(c) 2006 The Thomson Corp.

Set	Items	Description
S1	1702	DATA(3N) (WAREHOUS??? OR WARE()HOUS??? OR MART? ? OR REPOSI- TOR???) OR DATAMART? ? OR DATAWAREHOUS???
S2	22255	DATA(3N) (PREPROCESS??? OR PRE()PROCESS??? OR CLEAN??? OR C- LEANS??? OR PREPAR??? OR STANDARDIZ??? OR STANDARDIS??? OR ST- ANDARDIZATION OR STANDARDISATION OR STAGE? ? OR STAGING)
S3	366872	TRANSFORM??? OR TRANSFORMATION? ?
S4	772	KEY? ?(3N) (REUSE? ? OR REUSABLE OR REUSING OR (USE OR USES OR USED OR USING) (2N) (AGAIN OR OVER OR RE) OR DUPLICAT???)
S5	40600	DATE? ? OR DATESTAMP??? OR TIMESTAMP???
S6	200	S5(3N) (TRIGGER??? OR SCRAP??? OR HARVEST??? OR PULL OR PUL- LS OR PULLED OR PULLING OR LOG OR LOGS)
S7	3088	RDB OR RDBMS OR RELATIONAL() (DATABASE? ? OR DATA()BASE? ?)
S8	3	S1 AND S2(3N)S3
S9	0	S1 AND S2 AND S3 AND S4 AND S6
S10	1	S1 AND S2 AND S3 AND S4
S11	0	S10 NOT S8
S12	0	S1 AND S6
S13	1	S1 AND S4
S14	0	S13 NOT S8
S15	40	S1 AND S5
S16	3	S15 AND S7
S17	2	S16 NOT S8
S18	1	S15 AND S3
S19	0	S18 NOT (S8 OR S17)
S20	20	S15 NOT AD=20010208:20040208/PR
S21	15	S20 NOT AD=20040208:20060630/PR
S22	15	S21 NOT (S8 OR S17)

8/5/2 (Item 2 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2006 The Thomson Corp. All rts. reserv.

015269628 \*\*Image available\*\*

WPI Acc No: 2003-330557/200331

XRFX Acc No: N03-264662

**Subject-oriented data creation method for computer-controlled data  
warehousing system, involves preprocessing and transforming  
operational data output from source system and then stored in  
historical data warehouse**

Patent Assignee: OUTPUT TECHNOLOGY SOLUTIONS INC (OUTP-N)

Inventor: DUNHAM K; LECHER-MOORE R; RIGGS L T; SERIO T

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 20030014335	A1	20030116	US 2001780123	A	20010208	200331 B

Priority Applications (No Type Date): US 2001780123 A 20010208

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 20030014335	A1		24	G06F-017/60	

US 20030014335 A1 24 G06F-017/60

Abstract (Basic): US 20030014335 A1

NOVELTY - A processing unit (20) preprocesses operational data received from a source system and then transforms into subject-oriented data using reusable primary keys and RDBMS dates stored in an operating system of source system. The subject-oriented data is then stored in historical **data warehouse** (25).

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are included for the following:

- (1) subject-oriented data creation program; and
- (2) computer system for subject-oriented data creation.

USE - For computer-controlled **data warehousing** system.

ADVANTAGE - Permits user to easily and creatively use subject-oriented data for extraction and generation of useful information and reports.

DESCRIPTION OF DRAWING(S) - The figure shows the **data warehousing** system.

processing unit (20)

historical **data warehouse** (25)

pp; 24 DwgNo 2/7

Title Terms: SUBJECT; ORIENT; DATA; CREATION; METHOD; COMPUTER; CONTROL;  
DATA; WAREHOUSE; SYSTEM; TRANSFORM; OPERATE; DATA; OUTPUT; SOURCE; SYSTEM  
; STORAGE; HISTORY; DATA; WAREHOUSE

Derwent Class: T01

International Patent Class (Main): G06F-017/60

File Segment: EPI

22/5/11 (Item 2 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2006 The Thomson Corp. All rts. reserv.

016785741 \*\*Image available\*\*  
WPI Acc No: 2005-110017/200512  
Related WPI Acc No: 2005-345184  
XRPX Acc No: N05-094937

Check data extractor for bank, stores recognized payee name in check, in  
data warehouse if recognized name is present in list comprising names  
which are of particular interest to financial institution

Patent Assignee: NCR CORP (NATC )  
Inventor: HASSANEIN K S; MARLOW G G  
Number of Countries: 001 Number of Patents: 001  
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 6845366	B1	20050118	US 99444685	A	19991122	200512 B

Priority Applications (No Type Date): US 99444685 A 19991122

Patent Details:

Patent No	Kind	Lan Pg	Main IPC	Filing Notes
US 6845366	B1	12	G06F-015/30	

Abstract (Basic): US 6845366 B1

NOVELTY - A recognition engine recognizes the payee name in the payee field of a check, determined by processing the check image data. A memory stores a list comprising names which are of particular interest to a financial institution and which are changed selectively by the institution without any input from payees. The recognized payee name is stored in a **data warehouse**, if the name is present in the list.

USE - For extraction of data in courtesy amount, legal amount, payor, payee, **date** and signature fields of business and personal type check images received from check processing terminal e.g. automatic teller machine (ATM), bank teller station and check processing transport in image-based check processing system, and **date** and time of ATM transaction and identity of person using ATM card for ATM transaction, for storing in scalable **data warehouse** including non-volatile storage memory e.g. hard disk drive to be queried by financial institution e.g. bank for generating business-related information for providing customized marketing and customized bank statement, identification of major competitors in certain areas, and for providing information about geographic distribution of customers of bank's merchant customer for allowing merchant customers to customize advertisement/flyer campaigns towards certain neighborhoods.

ADVANTAGE - The useful information are extracted from the check easily.

DESCRIPTION OF DRAWING(S) - The figure shows a block diagram of the check data extractor.

pp; 12 DwgNo 2/8

Title Terms: CHECK; DATA; EXTRACT; BANK; STORAGE; RECOGNISE; NAME; CHECK;  
DATA; WAREHOUSE; RECOGNISE; NAME; PRESENT; LIST; COMPRISE; NAME; INTEREST  
; FINANCIAL; INSTITUTION

Derwent Class: T01; T05

International Patent Class (Main): G06F-015/30

File Segment: EPI

File 2:INSPEC 1898-2006/Jun W3  
(c) 2006 Institution of Electrical Engineers  
File 6:NTIS 1964-2006/Jun W3  
(c) 2006 NTIS, Intl Cpyrght All Rights Res  
File 8:Ei Compendex(R) 1970-2006/Jun W3  
(c) 2006 Elsevier Eng. Info. Inc.  
File 23:CSA Technology Research Database 1963-2006/Jun  
(c) 2006 CSA.  
File 34:SciSearch(R) Cited Ref Sci 1990-2006/Jun W4  
(c) 2006 Inst for Sci Info  
File 35:Dissertation Abs Online 1861-2006/Jun  
(c) 2006 ProQuest Info&Learning  
File 65:Inside Conferences 1993-2006/Jun 30  
(c) 2006 BLDSC all rts. reserv.  
File 94:JICST-EPlus 1985-2006/Mar W4  
(c)2006 Japan Science and Tech Corp(JST)  
File 95:TEME-Technology & Management 1989-2006/Jun W4  
(c) 2006 FIZ TECHNIK  
File 99:Wilson Appl. Sci & Tech Abs 1983-2006/May  
(c) 2006 The HW Wilson Co.  
File 111:TGG Natl.Newspaper Index(SM) 1979-2006/Jun 21  
(c) 2006 The Gale Group  
File 144:Pascal 1973-2006/Jun W1  
(c) 2006 INIST/CNRS  
File 239:Mathsci 1940-2006/Aug  
(c) 2006 American Mathematical Society  
File 256:TecInfoSource 82-2006/Aug  
(c) 2006 Info.Sources Inc  
File 434:SciSearch(R) Cited Ref Sci 1974-1989/Dec  
(c) 1998 Inst for Sci Info

Set	Items	Description
S1	19384	DATA(3N) (WAREHOUS??? OR WARE()HOUS??? OR MART? ? OR REPOSI-TOR???) OR DATAMART? ? OR DATAWAREHOUS???
S2	49772	DATA(3N) (PREPROCESS??? OR PRE()PROCESS??? OR CLEAN??? OR C-LEANS??? OR PREPAR??? OR STANDARDIZ??? OR STANDARDIS??? OR ST-ANDARDIZATION OR STANDARDISATION OR STAGE? ? OR STAGING)
S3	2653143	TRANSFORM??? OR TRANSFORMATION? ?
S4	847	KEY? ?(3N) (REUSE? ? OR REUSABLE OR REUSING OR (USE OR USES OR USED OR USING) (2N) (AGAIN OR OVER OR RE) OR DUPLICAT???)
S5	496635	DATE? ? OR DATESTAMP??? OR TIMESTAMP???
S6	3575	S5(3N) (TRIGGER??? OR SCRAP??? OR HARVEST??? OR PULL OR PUL-LS OR PULLED OR PULLING OR LOG OR LOGS)
S7	49321	RDB OR RDBMS OR RELATIONAL() (DATABASE? ? OR DATA()BASE? ?)
S8	124	S1 AND S2 AND S3
S9	39	S1 AND S2(3N)S3
S10	34	RD (unique items)
S11	18	S10 NOT PY=2002:2006
S12	0	S8 AND S4
S13	0	S8 AND S6
S14	3	S8 AND S5
S15	1	S1 AND S4
S16	0	S15 NOT PY=2002:2006
S17	1	S1 AND S6
S18	68	S1 AND S5 AND S7
S19	64	RD (unique items)
S20	1	S1 AND S5(5N)S7
S21	57	S19 NOT PY=2002:2006
S22	5	S21 AND S2:S3
S23	421	(DATE OR DATES OR DATESTAMP? ? OR TIMESTAMP? ?) (3N)EXTRACT-???

S24	0	S1 AND S23
S25	2994	S2 AND S3
S26	823	S25 AND DATA(3N)S3
S27	123	S26 AND (DATA()MINING OR OLAP)
S28	75	RD (unique items)
S29	25	S28 NOT PY=2002:2006
S30	56061	DATA()MINING OR OLAP
S31	6	S30 AND S4
S32	4	S30 AND S6
S33	3	S30 AND S5(5N)S7
S34	13	S31:S33
S35	8	RD (unique items)
S36	5	S35 NOT (S11 OR S14 OR S15 OR S17 OR S20 OR S22 OR S29)

11/5/3 (Item 3 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2006 Institution of Electrical Engineers. All rts. reserv.

07744353 INSPEC Abstract Number: C2000-12-6160D-006

**Title: A relational approach to data transformation**

Author(s): Shaffer, S.

Journal: DB2 Magazine vol.5, no.3 p.46-8, 50-4

Publisher: Miller Freeman,

Publication Date: Fall 2000 Country of Publication: USA

CODEN: DBMAF5

Material Identity Number: G132-2000-003

Language: English Document Type: Journal Paper (JP)

Treatment: Practical (P)

Abstract: Don't overlook SQL as your fundamental language for **cleansing** and transforming **data** for the **data warehouse**. Expert techniques will put you ahead of the game. (0 Refs)

Subfile: C

Descriptors: data handling; **data warehouses** ; relational databases; SQL

Identifiers: relational data transformation; SQL; **data warehouse** ; data cleansing

Class Codes: C6160D (Relational databases); C6130 (Data handling techniques)

Copyright 2000, IEE

11/5/5 (Item 5 from file: 2)

DIALOG(R) File 2:INSPEC

(c) 2006 Institution of Electrical Engineers. All rts. reserv.

07414926 INSPEC Abstract Number: C2000-01-6170K-007

**Title: Preprocessing and integration of data from multiple sources for knowledge discovery**

Author(s): Ceruti, M.G.; Kamel, M.N.

Author Affiliation: Space & Naval Warfare Syst. Center, San Diego, CA, USA

Journal: International Journal on Artificial Intelligence Tools (Architectures, Languages, Algorithms) vol.8, no.2 p.157-77

Publisher: World Scientific,

Publication Date: June 1999 Country of Publication: Singapore

CODEN: IAITEI ISSN: 0218-2130

SICI: 0218-2130(199906)8:2L.157:PIDF;1-Q

Material Identity Number: P897-1999-004

Language: English Document Type: Journal Paper (JP)

Treatment: Practical (P)

**Abstract:** The explosive growth in the generation and collection of data has generated an urgent need for a new generation of techniques and tools that can assist in transforming these data intelligently and automatically into useful knowledge. Knowledge discovery is an emerging multidisciplinary field that attempts to fulfill this need. Knowledge discovery is a large process that includes **data** selection, **cleaning**, **preprocessing**, integration, **transformation** and reduction, data mining, model selection, evaluation and interpretation, and finally consolidation and use of the extracted knowledge. This paper addresses the issues of data cleaning and integration for knowledge discovery by proposing a systematic approach for resolving semantic conflicts that are encountered during the integration of data from multiple sources. Illustrated with examples derived from military databases, the paper presents a heuristics-based algorithm for identifying and resolving semantic conflicts at different levels of information granularity. (16 Refs)

Subfile: C

Descriptors: computational linguistics; data analysis; data mining; knowledge acquisition

Identifiers: knowledge discovery; data preprocessing; data integration; multiple sources; useful knowledge; data selection; data mining; model selection; evaluation; interpretation; consolidation; extracted knowledge; data cleaning; semantic conflicts; military databases; heuristics-based algorithm; information granularity; database integration; command and control; **data warehousing**; semantic heterogeneity

Class Codes: C6170K (Knowledge engineering techniques); C6160 (Database management systems (DBMS)); C6180N (Natural language processing)

Copyright 1999, IEE



11/5/8 (Item 8 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2006 Institution of Electrical Engineers. All rts. reserv.

06655291 INSPEC Abstract Number: C9709-6130-008

**Title: Tools for traveling data [ warehouse databases]**

Author(s): Williams, J.

Journal: DBMS vol.10, no.7 p.69-70, 73-4, 76

Publisher: Miller Freeman,

Publication Date: June 1997 Country of Publication: USA

CODEN: DBMSEO ISSN: 1041-5173

SICI: 1041-5173(199706)10:7L:69:TTDW;1-W

Material Identity Number: M772-97008

Language: English Document Type: Journal Paper (JP)

Treatment: Practical (P)

Abstract: Loading **data** into a **warehouse** database is one of the most important steps in building a **data warehouse** . This article surveys the current tools for extracting, **cleaning** and **transforming data** , and analyzing and ensuring data quality. (0 Refs)

Subfile: C

Descriptors: data handling; quality control; very large databases

Identifiers: data loading; warehouse database; **data warehouse** ; data extraction; data cleaning; data transformation; data quality analysis; data handling tools; data migration

Class Codes: C6130 (Data handling techniques); C6160Z (Other DBMS)

Copyright 1997, IEE

11/5/9 (Item 9 from file: 2)  
DIALOG(R)File 2:INSPEC  
(c) 2006 Institution of Electrical Engineers. All rts. reserv.

06607206 INSPEC Abstract Number: C9707-6160Z-009

**Title:** Data warehousing : cleaning **and** transforming data

**Author(s):** White, C.

**Journal:** Info DB vol.10, no.6 p.11-12

**Publisher:** Database Associates Int,

**Publication Date:** April 1997 **Country of Publication:** USA

**CODEN:** IFDBEB **ISSN:** 0891-6004

**SICI:** 0891-6004(199704)10:6L.11:DWCT;1-T

**Material Identity Number:** L966-97004

**Language:** English **Document Type:** Journal Paper (JP)

**Treatment:** Practical (P)

**Abstract:** One of the more difficult tasks in building a **data warehouse** is **cleaning** and **transforming** source **data** . This article discusses a data transformation management system (DTMS) architecture that can be used to implement a **data cleanup** and **transformation** project. Although a DTMS can be used for many different types of **data cleanup** and **transformation** , the focus of this article is on the use of a DTMS for building a **data warehouse**. (0 Refs)

**Subfile:** C

**Descriptors:** business data processing; data analysis; data integrity;  
very large databases

**Identifiers:** **data warehouse** ; source data cleaning; DTMS architecture;  
data transformation management system; data cleanup

**Class Codes:** C6160Z (Other DBMS); C6130 (Data handling techniques);  
C7100 (Business and administration)

Copyright 1997, IEE

17/5/1 (Item 1 from file: 95)  
DIALOG(R)File 95:TEME-Technology & Management  
(c) 2006 FIZ TECHNIK. All rts. reserv.

01441718 20000802168

**Spurensuche auf elektronischen Pfaden. E-Intelligence: Wirtschaftlichkeit messen**

Doerken, J  
Infoware, D  
Computerwoche Extra, v27, n5, pp52-55, 2000  
Document type: journal article Language: German  
Record type: Abstract  
ISSN: 0935-1310

**ABSTRACT:**

Die Chancen, die sich Firmen durch E-Commerce bieten, sind mittlerweile allgemein bekannt. Dem E-Commerce werden fuer die naechsten Jahre gewaltige Umsatz-Steigerungen vorausgesagt. Dennoch ist es fuer jeden einzelnen Anbieter im Internet wichtig zu wissen, wie Kunden und Interessenten das bereitgehaltene Web-Angebot nutzen. Nur so kann man das Potenzial, welches das Internet bietet, Gewinn bringend nutzen. Hierzu bieten sich sogenannte E-Intelligence-Tools an. Waren es frueher reine Informationen ueber angebotene Produkte oder Dienstleistungen, so hat sich dies eindeutig Richtung Handel gewandelt. In dieser Umbruchsphase ging bei vielen Firmen die erforderliche Wirtschaftlichkeitsrechnung des E-Commerce unter. Das entscheidende Kriterium hierfuer ist die Kundenzufriedenheit, die durch das im Internet Angebotene erreicht wird. Diese laesst sich jedoch nicht einfach durch die Anzahl der taeglichen Hits feststellen. Vielmehr sind Verweildauer, typische Pfade oder Ein- und Ausstiegspunkte von Bedeutung. Basis hierfuer sind Logfiles, die beim Zugriff auf die Web-Seiten erstellt werden. Bei stark frequentierten Seiten muss aufgrund der Datenmenge dafuer gesorgt werden, dass die erhaltenen **Log - Daten** automatisch in das **Datawarehouse** importiert werden. Diese gewonnenen Daten koennen ggf. mit Kundendaten verknuepft werden, um Data-Mining-Analysen durchzufuehren. Da Ergebnis sind im Idealfall Kundenprofile. E-Intelligence in seiner Gesamtheit zeichnet sich daher durch folgendes aus: Es liefert Erkenntnisse, wie Web-Besucher die Seiten nutzen, es ermoeeglicht Kundensegmentierung, es hilft, die profitabelsten Kundenkreise zu identifizieren, es kann die Interessen von Kunden voraussagen, es definiert und bewertet die Bedeutung des E-Commerce, es liefert Grundlagen fuer das richtige Web-Design, es bringt die richtigen Kunden immer wieder zurueck, und ausserdem ueberwacht und optimiert es die technische Performance.

DESCRIPTORS: VALIDATION METHOD; BENEFIT COST ANALYSIS; COST OPTIMIZATION; MARKET ANALYSIS; RENTABILITY; STATISTICAL CONTROL; SYSTEM OPTIMIZATION; DEVELOPMENTAL TREND; CONSUMERS BEHAVIOUR; EFFICIENCY--PROFITABILITY; WORLD WIDE WEB; FUTURE DEMAND; ELECTRONIC COMMERCE; CLIENTS  
IDENTIFIERS: E INTELLIGENCE; **LOG DATEN** ; KUNDENZUFRIEDENHEIT;  
E-Intelligence; Wirtschaftlichkeit; Erfolgskontrolle

## Tracing on electronic paths. E-Intelligence: Economy measure

**ABSTRACT:** The chances, which are offered to companies by E-Commerce, are meanwhile well-known. The E-Commerce for the next years enormous increases in sales are forecast. It nevertheless is important for each individual offerer in the Internet to know, how customers and prospective customers use the held ready Web offer. Only so one can use the Potenzial, which Internet offers, profit bringing. For this so-called E-Intelligence-Tools offers itself. If there was in former times pure information over offered products or services, then this changed itself clearly direction trade. Regarding this paging phase the necessary economy calculation of the E-Commerce went down at many companies. The crucial criterion for this is the customer satisfaction, which is reached by in the Internet the offering. This does not let itself determine however simply by the number of daily hits. Rather are period spent, typical paths or in and points of door of importance. Basis for this are log files, which are provided with the access to the web pages. With strongly frequented sides it must be ensured due to the data set that the received logs - data into the Datawarehouse to be imported automatically. These won data can be linked if necessary with customer data, in order to accomplish DATA Mining analyses. There result are ideally customer profiles. E-Intelligence in its Whole is characterised therefore by the following: It supplies Realizations, how Web visitors use the sides, makes possible Customer segmenting, it helps, the profitabelsten clientele too identify, it can the interests by customers forecast, it defined and evaluate the meaning of the E-Commerce, it supply bases for the correct Web Design, it return the correct customers again and again, and in addition supervised and optimize it the technical performance.

translation on 6/30/2006 from:

[http://www.google.com/language\\_tools?hl=en](http://www.google.com/language_tools?hl=en)

22/5/4 (Item 4 from file: 95)  
DIALOG(R)File 95:TEME-Technology & Management  
(c) 2006 FIZ TECHNIK. All rts. reserv.

01208038 E98060166363

**Tools behindern anfangs eher. Extraktions- und Transformations -Werkzeuge  
- (k)ein Silberstreifen am Horizont?**

Hebestreit, U  
Saphir Unternehmensberatung, Muenster, D  
Computerwoche Focus, v37, n2, pp22-23, 1998  
Document type: journal article Language: German  
Record type: Abstract  
ISSN: 0935-1329

**ABSTRACT:**

Die Ausfuehrungen behandeln Probleme, die in Zusammenhang mit der Anwendung spezieller Werkzeuge beim Laden und dem Betrieb eines **Data Warehouse** mit den geeigneten Informationsdaten auftreten. Derartige Werkzeuge koennen in der Regel etwa 80 Prozent der Aufgabenstellung erfuellen, waehrend die restlichen 20 Prozent nur schwer oder gar nicht in dem vom Werkzeug gesetzten Rahmen loesbar sind. Ein zentrales Problem der Bewirtschaftung besteht z.B. darin, Aenderungen in den **Daten** seit der letzten Extraktion zu erkennen, hierfuer koennen u.a. die nativen Replikationstools der jeweiligen **RDEMS** -Hersteller zum Einsatz kommen. Ein weiterer zu beachtender Aspekt besteht in der logischen und physischen Trennung der Extraktions- und Transformationsprozesse, um vor allem Datenausfaelle zu vermeiden. Weitere Problemkreise beinhalten u.a. Aspekte der Zusammenfuehrung von Attributen aus verschiedenen Quellsystemen zu einer Entitaet im **Data Warehouse**, Fragen von Hierarchien und sogenannten 'slowly changing dimensions', das heisst solchen Objekten (oder ihrer Beziehungen untereinander), die sich im Laufe der Zeit nur selten aendern. Wesentliche Bedeutung fuer die Bewirtschaftung des **Data Warehouse** besitzt auch die Automatisierung des Gesamtprozesses, neben dem bereits erwaehnten Gesichtspunkt der Parametrisierung spielen hier formale und inhaltliche Gesichtspunkte eine entscheidende Rolle. Die Anwendung geeigneter Tools bei der Bewirtschaftung eines **Data Warehouse** kann zumindest anfangs die Arbeit eher behindern als unterstuetzen, Vorteile wie einheitlicher Programmierstil und einheitlicher Zugriff auf Datenstrukturen der beteiligten Systeme und 'Metadaten' treten oftmals erst in spaeteren Phasen zutage, ganz in Abhaengigkeit der Lernkurve und der Loesung komplexer Probleme.

DESCRIPTORS: COMPUTER PROCESSING; SOFTWARE TOOLS; INFORMATION PRESENTATION;  
INFORMATION FLOW; INFORMATION CONTENT; INFORMATION MANAGEMENT; INFORMATION  
RETRIEVAL SYSTEMS; INFORMATION SYSTEMS; INFORMATION TECHNOLOGY  
IDENTIFIERS: **DATA WAREHOUSE** ; **Data Warehouse** ; Software-Werkzeug

Tools obstruct at first rather. Extraction and transformation - tools - (k) a silver strip on the horizon?

**ABSTRACT:** The remarks treat problems, which arise in connection with the application of special tools with loading and the enterprise DATA of a Warehouse with the suitable information data. Such tools can fulfill usually about 80 per cent of setting of tasks, while the remaining 20 per cent only heavily or not at all in the framework set by the tool are solvable. A central problem that management consists e.g. of recognizing changes in the data since the last extraction for this can among other things the native Replikationstools of the respective RDBMS - manufacturers be used. A further too considering aspect exists in the logical and physical separation of the extraction and transformation processes, in order to avoid above all data losses. Further problem areas contain among other things aspects that Unification of attributes from different source systems to a Entitaet in DATA the Warehouse, questions of hierarchies and so-called "slowly chanching dimension", i.e. such objects (or their relations among themselves), which change only rarely in the course of the time. Substantial meaning for the management DATA of the Warehouse possesses also the automation of the total process, apart from the criterion of parameterizing already mentioned plays here formal and contentwise criteria a crucial role. Application suitable Tools with that management DATA of a Warehouse can obstruct at least at the beginning of the work rather as support, advantages like uniform programming style and uniform accesses to data structures of the systems and "meta data involved" steps often only into later phases to light, completely in dependence of the learning curve and the solution complex problems.

translation on 6/30/2006 from:

[http://www.google.com/language\\_tools?hl=en](http://www.google.com/language_tools?hl=en)

29/5/10 (Item 10 from file: 2)

DIALOG(R) File 2:INSPEC

(c) 2006 Institution of Electrical Engineers. All rts. reserv.

07414926 INSPEC Abstract Number: C2000-01-6170K-007

**Title:** Preprocessing and integration of data from multiple sources for knowledge discovery

Author(s): Ceruti, M.G.; Kamel, M.N.

Author Affiliation: Space & Naval Warfare Syst. Center, San Diego, CA, USA

Journal: International Journal on Artificial Intelligence Tools (Architectures, Languages, Algorithms) vol.8, no.2 p.157-77

Publisher: World Scientific,

Publication Date: June 1999 Country of Publication: Singapore

CODEN: IAITEI ISSN: 0218-2130

SICI: 0218-2130(199906)8:2L.157:PIDF;1-Q

Material Identity Number: P897-1999-004

Language: English Document Type: Journal Paper (JP)

Treatment: Practical (P)

**Abstract:** The explosive growth in the generation and collection of data has generated an urgent need for a new generation of techniques and tools that can assist in **transforming** these **data** intelligently and automatically into useful knowledge. Knowledge discovery is an emerging multidisciplinary field that attempts to fulfill this need. Knowledge discovery is a large process that includes **data** selection, **cleaning**, **preprocessing**, integration, **transformation** and reduction, **data mining**, model selection, evaluation and interpretation, and finally consolidation and use of the extracted knowledge. This paper addresses the issues of **data cleaning** and integration for knowledge discovery by proposing a systematic approach for resolving semantic conflicts that are encountered during the integration of data from multiple sources. Illustrated with examples derived from military databases, the paper presents a heuristics-based algorithm for identifying and resolving semantic conflicts at different levels of information granularity. (16 Refs)

Subfile: C

Descriptors: computational linguistics; data analysis; **data mining**; knowledge acquisition

Identifiers: knowledge discovery; **data preprocessing**; data integration; multiple sources; useful knowledge; data selection; **data mining**; model selection; evaluation; interpretation; consolidation; extracted knowledge; **data cleaning**; semantic conflicts; military databases; heuristics-based algorithm; information granularity; database integration; command and control; data warehousing; semantic heterogeneity

Class Codes: C6170K (Knowledge engineering techniques); C6160 (Database management systems (DBMS)); C6180N (Natural language processing)

Copyright 1999, IEE

29/5/12 (Item 12 from file: 2)  
DIALOG(R)File 2:INSPEC  
(c) 2006 Institution of Electrical Engineers. All rts. reserv.

07164271 INSPEC Abstract Number: C1999-03-6160Z-011

**Title:** I/O problems in preparing data for data warehousing and data mining . 1

**Author(s):** Won Kim

**Author Affiliation:** Cyber Database Solutions, Austin, TX, USA

**Journal:** JOOP vol.11, no.9 p.13-14, 17

**Publisher:** SIGS Publications,

**Publication Date:** Feb. 1999 **Country of Publication:** USA

**CODEN:** JOOPEC **ISSN:** 0896-8438

**SICI:** 0896-8438(199902)11:9L.13:PPDD;1-Z

**Material Identity Number:** G316-1999-002

**Language:** English **Document Type:** Journal Paper (JP)

**Treatment:** Practical (P)

**Abstract:** The "data preparation" step is one of the most time-consuming and critically important steps in both the data warehousing and **data mining** process. Before a target data warehouse (or data mart) is loaded, the data extracted from various data sources must in general undergo a series of **transformations**. Before a **data** set or a data warehouse may be fed to a **data mining** algorithm, it must in general also undergo a series of **transformations**. The **data** preparation steps in data warehousing and **data mining** have many types of **transformations** in common. Although relational database systems (RDBs) are increasingly being adopted to store data warehouses and data sets for **data mining** (and database marketing applications), they are not well-suited for many of these **transformations**. RDBs have largely been designed to support online transaction processing (OLTP) applications that tend to retrieve and update a small number of records from a large database based on a specific combination of search conditions. RDBs resort to indexing or hashing techniques to zoom in quickly on the records that satisfy the search conditions. This article examines the data preparation steps in data warehousing and **data mining**, with a view to eliciting the processing requirements on database systems. (2 Refs)

**Subfile:** C

**Descriptors:** **data mining**; data preparation; data warehouses; relational databases; transaction processing

**Identifiers:** I/O problems; data preparation; data warehousing; **data mining**; data marts; **data transformations**; relational database systems; data sets; database marketing applications; online transaction processing; record retrieval; record updating; search conditions; indexing techniques; hashing techniques; processing requirements; client/server systems

**Class Codes:** C6160Z (Other DBMS); C6160D (Relational databases); C6130 (Data handling techniques)

Copyright 1999, IEE



29/5/21 (Item 2 from file: 144)  
DIALOG(R)File 144:Pascal  
(c) 2006 INIST/CNRS. All rts. reserv.

15292137 PASCAL No.: 01-0464704

**Mining E-commerce data: The good, the bad, and the ugly**

**Advances in knowledge discovery and data mining : Hong Kong, 16-18**

**April 2001**

KOHAVI Ronny

CHEUNG David, ed; WILLIAMS Graham J, ed; QING LI, ed

PAKDD 2001 : Pacific-Asia conference on advances in knowledge discovery  
and data mining, 5 (Hong Kong CHN) 2001-04-16

Journal: Lecture notes in computer science, 2001, 2035 p. 2

ISBN: 3-540-41910-1 ISSN: 0302-9743 Availability: INIST-16343;

354000092387270020

Document Type: P (Serial); C (Conference Proceedings); E (Summary) ; A (Analytic)

Country of Publication: Germany; United States

Language: English

Electronic commerce provides all the right ingredients for successful **data mining** (the Good). Web logs, however, are at a very low granularity level, and attempts to mine e-commerce data using only web logs often result in little interesting insight (the Bad). Getting the data into minable formats requires significant **pre**-processing and **data transformations** (the Ugly). In the ideal e-commerce architecture, high level events are logged, **transformations** are automated, and **data mining** results can easily be understood by business people who can take action quickly and efficiently. Lessons, stories, and challenges based on mining real data at Blue Martini Software will be presented.

English Descriptors: Electronic trade; Data processing; Pretreatment;  
Database; **Data mining**

French Descriptors: Commerce electronique; Traitement donnee; Pretraitement  
; Base donnee; Fouille donnee

Classification Codes: 001D04B03; 001D02B07D

Copyright (c) 2001 INIST-CNRS. All rights reserved.

36/5/4 (Item 4 from file: 2)

DIALOG(R) File 2:INSPEC

(c) 2006 Institution of Electrical Engineers. All rts. reserv.

07900117 INSPEC Abstract Number: C2001-05-6160D-006

**Title: Mining relational databases**

Author(s): Moal, F.; Turmeaux, T.; Vrain, C.

Author Affiliation: Orleans Univ., France

Conference Title: Principles of Data Mining and Knowledge Discovery. 4th European Conference, PKDD 2000. Proceedings (Lecture Notes in Artificial Intelligence Vol.1910) p.536-41

Editor(s): Zighed, D.A.; Komorowski, J.; Zytchow, J.

Publisher: Springer-Verlag, Berlin, Germany

Publication Date: 2000 Country of Publication: Germany xv+701 pp.

ISBN: 3 540 41066 X Material Identity Number: XX-2000-02602

Conference Title: Principles of Data Mining and Knowledge Discovery. 4th European Conference, PKDD 2000

Conference Date: 13-16 Sept. 2000 Conference Location: Lyon, France

Language: English Document Type: Conference Paper (PA)

Treatment: Practical (P); Theoretical (T)

Abstract: We propose a classification system to induce an intentional definition of a relation from examples, when background knowledge is stored in a relational database composed of several tables and views. Refinement operators have been defined to integrate in a uniform way different induction tools learning numeric and symbolic constraints. The particularity of our approach is to **use** integrity constraints **over** the database ( **keys** and foreign keys) to explore the hypotheses space. Moreover new attributes can be introduced, relying on the aggregation operator "group by". (8 Refs)

Subfile: C

Descriptors: data integrity; **data mining** ; database theory; learning by example; relational algebra; relational databases; very large databases

Identifiers: relational database mining; classification system; intentional definition; tables; views; refinement operators; learning by example; integrity constraints; hypotheses space; aggregation operator; large databases; **data mining** ; relational algebra

Class Codes: C6160D (Relational databases); C6170K (Knowledge engineering techniques); C1230L (Learning in AI); C6160Z (Other DBMS); C4210 (Formal logic); C4250 (Database theory)

Copyright 2001, IEE

File 348:EUROPEAN PATENTS 1978-2006/ 200626  
(c) 2006 European Patent Office  
File 349:PCT FULLTEXT 1979-2006/UB=20060622,UT=20060615  
(c) 2006 WIPO/Univentio

Set	Items	Description
S1	5305	DATA(3N) (WAREHOUS??? OR WARE()HOUS??? OR MART? ? OR REPOSI- TOR???) OR DATAMART? ? OR DATAWAREHOUS???
S2	28051	DATA(3N) (PREPROCESS??? OR PRE()PROCESS??? OR CLEAN??? OR C- LEANS??? OR PREPAR??? OR STANDARDIZ??? OR STANDARDIS??? OR ST- ANDARDIZATION OR STANDARDISATION OR STAGE? ? OR STAGING)
S3	311809	TRANSFORM??? OR TRANSFORMATION? ?
S4	1576	KEY? ?(3N) (REUSE? ? OR REUSABLE OR REUSING OR (USE OR USES OR USED OR USING) (2N) (AGAIN OR OVER OR RE) OR DUPLICAT???)
S5	2206429	DATE? ? OR DATESTAMP??? OR TIMESTAMP???
S6	1348	S5(3N) (TRIGGER??? OR SCRAP??? OR HARVEST??? OR PULL OR PUL- LS OR PULLED OR PULLING OR LOG OR LOGS)
S7	8456	RDB OR RDBMS OR RELATIONAL() (DATABASE? ? OR DATA()BASE? ?)
S8	67	S1(100N)S2(5N)S3
S9	0	S8(100N)S4
S10	7	S8(100N)S5
S11	0	S1(100N)S6(10N)S7
S12	4	S1(100N)S6
S13	4	S12 NOT S10
S14	11	S1(100N)S4
S15	11	S14 NOT (S10 OR S13)
S16	33	S1(100N)S5(10N)S7
S17	33	S16 NOT (S10 OR S13 OR S15)
S18	18	S17 NOT AD=20010208:20040208/PR
S19	15	S18 NOT AD=20040208:20060630/PR
S20	15	S19 AND IC=G06F
S21	60	S8 NOT (S10 OR S13 OR S15 OR S17)
S22	27	S21 NOT AD=20010208:20040208/PR
S23	21	S22 NOT AD=20040208:20060630/PR
S24	19	S23 AND IC=G06F

10/3,K/5 (Item 4 from file: 349)  
DIALOG(R)File 349:PCT FULLTEXT  
(c) 2006 WIPO/Univentio. All rts. reserv.

00867313 \*\*Image available\*\*

**COMPUTER METHOD AND DEVICE FOR TRANSPORTING DATA**  
**PROCEDE ET DISPOSITIF INFORMATIQUE SERVANT A ACHEMINER DES DONNEES**

Patent Applicant/Assignee:

INFORMATICA CORPORATION, 1200 Chrysler Drive, Menlo Park, CA 94025, US,  
US (Residence), US (Nationality)

Inventor(s):

KANCHWALLA Firoz, 396 Ano Nuevo Avenue, Apt. #310, Sunnyvale, CA 94085,  
US,  
LYLE David, 331 Los Gatos Boulevard, Los Gatos, CA 95032, US,  
BAIS Sujit, 655 South Fair Oaks Avenue, Apt. #L107, Sunnyvale, CA 94086,  
US,  
MAADAPUSI Srinivasan, 450 North Mathilda Avenue, Apt. #E202, Sunnyvale,  
CA 94085, US,  
DONGRE Amol, 655 South Fair Oaks Avenue, Apt. #I-107, Sunnyvale, CA 94086  
, US,  
SOMAKUMAR Premkumar, 655 South Fair Oaks Avenue, Apt. #C-108, Sunnyvale,  
CA 94086, US,

Legal Representative:

GALLENSON Mavis S (et al) (agent), 5670 Wilshire Blvd. Suite 2100, Los  
Angeles, CA 90036, US,

Patent and Priority Information (Country, Number, Date):

Patent: WO 200201415 A2-A3 20020103 (WO 0201415)  
Application: WO 2001US19768 20010621 (PCT/WO US0119768)  
Priority Application: US 2000214299 20000626; US 2001877370 20010607

Designated States:

(Protection type is "patent" unless otherwise stated - for applications  
prior to 2004)

AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ  
EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR  
LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL  
TJ TM TR TT TZ UA UG UZ VN YU ZA ZW  
(EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE TR  
(OA) BF BJ CF CG CI CM GA GN GW ML MR NE SN TD TG  
(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZW  
(EA) AM AZ BY KG KZ MD RU TJ TM

Publication Language: English

Filing Language: English

Fulltext Word Count: 7109

Fulltext Availability:

Detailed Description

Detailed Description

... required for loading the database, allowing

13

for a smaller operational window. In addition, the - **staging** area  
**prepares** the **data** consistently for loading into the analytic data  
interface from various sources.

Continuing with Figure 3...

...and their

interrelationship as required by the analytic data interface 5.

Analytic data interface 5 **transforms** data for loading into **data**  
**warehouse** 6 for use in applications 8. In the present embodiment, the  
analytic **data** interface **cleans** **data** by enforcing commonalties in

**dates** ,  
names and other data types that appear across multiple systems and  
prepares it for the source-independent **data warehouse** .

In the present embodiment, analytic data interface 5 includes a  
graphical user interface that makes...  
...and customize how business data is loaded into an analytic applications  
system such as a **data warehouse** 6. Analytic **data** interface 5  
includes a simplified abstraction layer  
1 4  
for the **data warehouse** administrator, allowing the warehouse  
administrator to configure how data is loaded into the analytic  
applications...